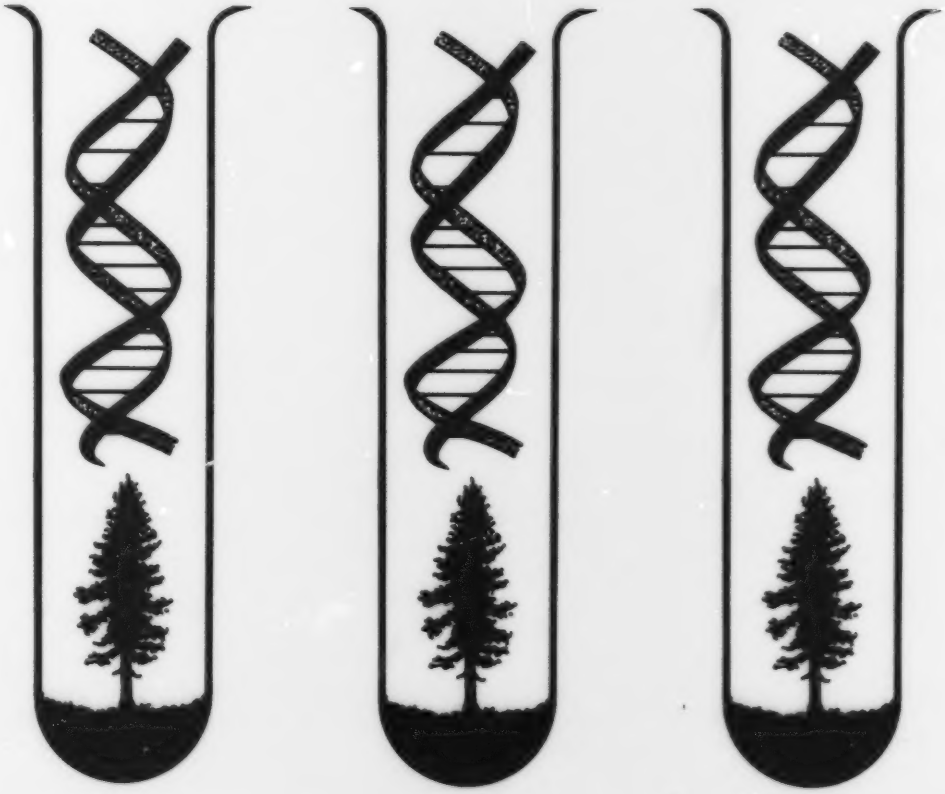
Forestry Canada Forêts Canada

Constructing conifer genomic libraries: a basic guide

L.L. DeVerno, J.R. Byrne, J.A. Pitel, and W.M. Cheliak

Information Report PI-X-88
Petawawa National Forestry Institute



PETAWAWA NATIONAL FORESTRY INSTITUTE

In common with the rest of Forestry Canada, the Petawawa National Forestry Institute has as its objective the promotion of better management and wiser use of Canada's forest resource to the economic and social benefit of all Canadians. Objectives of program activities carried out at the Institute support this goal through discovery, development, demonstration, implementation, and transfer of innovations. Because it is a national institute, particular emphasis is placed on problems that transcend regional boundaries or that require special expertise and equipment that cannot be duplicated in Forestry Canada regional establishments. Such research is often performed in close cooperation with staff of the regional centres, provincial forest services, and the forest industry.

Research initiatives and technical services at the Institute encompass five major activities:

FOREST GENETICS AND BIOTECHNOLOGY — Integrates projects in tree genetics, soil microbiology, micropropagation, molecular genetics, meteorology, and seed research. It also includes the client services and seed bank operations of the National Tree Seed Centre, a long-standing program with extensive international affiliations.

FOREST MANAGEMENT SYSTEMS — This program integrates projects in fire, remote sensing, modelling, growth and yield, and forest pest management to provide research and development for the formulation and demonstration of forest management systems.

NATIONAL FOREST RESOURCE STATISTICS — Provides biological, technical, and socioeconomic information on Canada's forest-based resources. The program involves progressive development of databases and establishment of new databases and software in support of policy development in forestry. The Forest Inventory Program collates information on the forest resource at a national level, maintains the Canadian Forest Resources Data System, and prepares the national forest inventory.

COMMUNICATIONS — Integrates activities of the library, public awareness, information, and editing and publications projects. The Institute is visited by more than 20 000 people every year. There is a Visitor Centre for the public, self-guided tours, and an extensive education project. The national repository of all scientific and technical publications of the Forestry Canada and the principal Forestry Canada publications distribution centre are both located at PNFI.

THE RESEARCH FOREST — Besides natural stands manipulated in a variety of ways for silvicultural research, the 100 km^2 Petawawa Forest contains extensive areas of plantations dating back six decades. Research plantations are a source of growth and yield data derived from cultural experiments, and they are becoming valuable for pedigreed genetic materials for micropropagation and molecular genetics studies. The forest also offers opportunities for short- and long-term testing of forest management strategies.

CONSTRUCTING CONIFER GENOMIC LIBRARIES: A BASIC GUIDE

L.L. DeVerno, J.R. Byrne, J.A. Pitel, and W.M. Cheliak

Information Report PI-X-88
Petawawa National Forestry Institute
Forestry Canada
1989

90-00223/

©Minister of Supply and Services Canada 1989
Catalogue No. Fo46-11/88-1989E
ISBN 0-662-17012-1
ISSN 0706-1854
Printed in Canada

Copies of this publication may be obtained free of charge from the following address:

Forestry Canada
Publications Distribution Centre
Petawawa National Forestry Institute
Chalk River, Ontario
K0J 1J0

Telephone: 613-589-2880

A microfiche edition of this publication may be purchased from:

Micromedia Ltd.
Place du Portage
165, Hôtel-de-Ville
Hull, Québec
J8X 3X2

Cette publication est également disponible en français sous le titre
Création d'une banque génomique pour les conifères : Un guide de base.

Table of Contents

v Abstract

1 1. Introduction

4 2. Isolation of high molecular weight conifer DNA

4 2.1 DNA isolation using modified CTAB procedure
5 2.1.1 Precautions
6 2.1.2 Troubleshooting
7 2.2 Genomic DNA isolation from tissue-cultured plant cells

8 3. Fractionation of donor DNA

8 3.1 Preparation of 20 kb fragments of eucaryotic DNA
8 3.1.1 Fractionation of donor DNA
9 3.1.2 Troubleshooting

10 4. Vector preparation

10 4.1 Purification of intact phage
12 4.2 Extraction and purification of phage DNA
12 4.3 Fractionation of vector DNA and purification of phage arms
12 4.3.1 Cleavage of vector DNA
12 4.3.2 Treatment of fractionated vector DNA with alkaline phosphatase
13 4.3.3 Troubleshooting

14 5. Significance of host cell strain when cloning plant genomic DNA

15 6. Ligation, packaging, and titering

15 6.1.1 Determination of optimal ligation conditions
16 6.1.2 Troubleshooting
17 6.2 Large scale ligation and packaging

18 7. Amplificiation and storage

18 8. Evaluation of genomic DNA library quality

18 9. Concluding remarks

19 References

20 Acknowledgments

21 Appendix I: Methods for DNA purificiation
23 Appendix II: Media and buffer formulations
26 Appendix III: Table 1: Number of recombinants to represent genomes
26 Appendix IV: Table 2: 5-methylcytosine in plant DNA

Figures

2 1. Constructing a library of size-fractionated conifer tissue DNA
11 2. Simplified map of the EMBL3 vector

ABSTRACT

Procedures are given for constructing genomic DNA bacteriophage libraries from conifer tissue. Although black spruce *Picea mariana* (Mill.) B.S.P. tissue was used, the protocols can be applied to any plant species. Included are techniques for high molecular weight genomic DNA isolation from mature needles, seedlings, and cell suspension cultures. In addition, purification of size-fractionated genomic DNA fragments, vector DNA isolation, and preparation of vector arms are outlined. Details are provided for production of recombinant clones, in vitro packaging, amplification, and evaluation of library quality. Appendices contain methods for purification of sized fragments, media and buffer formulas, and the number of recombinants required to represent genomes of various species.

RÉSUMÉ

Ce rapport décrit la manière de créer une banque de bactériophages à ADN génomique à partir de tissus de conifères, en l'occurrence des épinettes noires (*Picea mariana* [Mill.] B.S.P.). La méthode peut être appliquée à toutes les espèces végétales. On présente notamment des techniques pour isoler des fragments d'ADN génomique à poids moléculaire élevé à partir d'aiguilles à maturité, de semis et de cultures contenant des cellules en suspension. Les auteurs décrivent également la purification de fragments d'ADN génomique fractionnés en fonction de la taille, l'isolation de l'ADN vecteur et la préparation des bras de vecteur. Y sont également détaillés la production de clones recombinants, l'emballage *in vitro*, l'amplification et l'évaluation de la qualité de la banque. Les annexes contiennent des méthodes de purification des fragments, des milieux et des préparations de tampons. Elles font aussi état du nombre de recombinants nécessaires pour représenter les génomes de diverses espèces.

CONSTRUCTING CONIFER GENOMIC LIBRARIES: A BASIC GUIDE

1. INTRODUCTION

Biotechnology offers an innovative approach to assist traditional tree breeding programs. However, a foundation of basic research must be established before this technology can be applied. Biotechnological research on forest tree species is focused currently on three major areas:

1. Genetic transformation – Many different techniques can be used to introduce desirable genes into host cells. These include microinjection, use of microprojectiles, electrical or chemical poration, and use of the crown gall bacterium *Agrobacterium tumefaciens*.
2. Plant cell and tissue culture – This is one method to produce, maintain, and multiply viable plantlets containing desirable genes and/or gene combinations obtained by either transformation or traditional selection methods.
3. Molecular biology – Molecular biology techniques are used to identify and characterize genes. Elements controlling expression of these genes need to be understood before transfer of specific genes or gene combinations can produce predictable phenotypes.

There are two types of DNA libraries, cDNA and genomic, that can be constructed to assist in identifying and isolating genes. Ideally, a cDNA (complementary DNA) library contains sequences reflecting all mRNA sequences transcribed in a cell. Thus, a cDNA library facilitates the study of gene structure, organization, and expression of genes that are biologically active in a specific tissue and/or during the developmental stage when the mRNA was isolated. By contrast, a genomic DNA library is a collection of randomly cloned fragments that, when considered as a whole, theoretically contain all sequences from a given genome. Cloned fragments are large enough to contain whole genes, linked genes, and their flanking sequences in stable form.

Preparing a genomic DNA library involves cleavage of genomic DNA with a restriction enzyme to produce sized fragments, which can then be cloned into a chosen vector. The fraction of DNA that can be incorporated in the gene library (the clonable fraction) is that portion of the fragments which lies between the minimum and maximum allowable insert size for the chosen carrier vector. However, this assumes every internucleotide bond has the same probability of being the location of a restriction enzyme cleavage site, and that rates of cleavage do not change from site to site in the DNA to be cloned. If each clonable DNA fragment is of identical size, and if the genome size is known, then the size of the library that theoretically will have any probability of including a specific DNA sequence is described by the formula (Clarke and Carbon 1976):

$$N=\ln(1-p)/\ln(1-f)$$

where p is the probability of containing any particular DNA sequence, f is the fraction of the genome represented by each fragment, and N corresponds to the number of cloned fragments and, therefore, the necessary number of recombinant clones. Appendix III presents, for various species, the number of recombinants required to achieve 99% probability (p=0.99) of having a given DNA sequence represented in a library of 17 kb fragments.

A theoretically complete library (i.e. one that contains the appropriate number of recombinants at p=0.99) may lack particular genomic sequences or underrepresent others. Distribution of cleavage sites will affect representation of sequences. As well, particular regions of the eucaryotic chromosome may contain certain DNA sequences that are detrimental to growth of bacteriophage λ and may be lost from the library. Recombination events during phage propagation can lead to unwanted deletions or alterations of cloned sequences.

The basic steps used in constructing genomic DNA libraries are shown in Figure 1. They are, in summary:

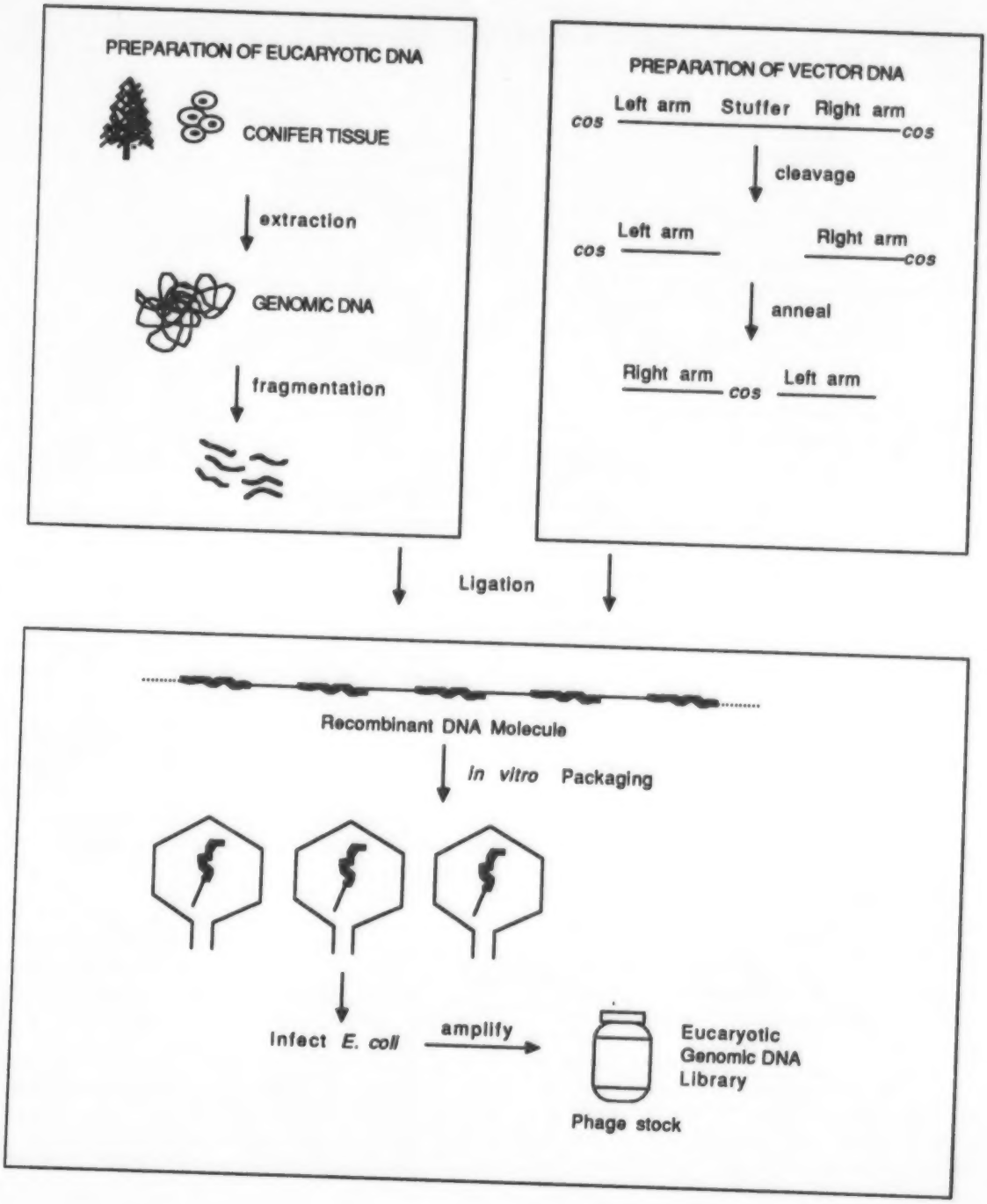


Figure 1. Constructing a library of size-fractionated conifer tissue DNA

(1) isolation and purification of eucaryotic DNA
(2) cleavage of DNA into appropriate sizes for cloning
(3) preparation of vector DNA
(4) ligation of sized eucaryotic DNA with vector DNA
(5) packaging recombinant molecules, titering
(6) amplification and storage of clones, and
(7) analysis

The aim of this report is to outline, in an organized manner, the production of genomic bacteriophage libraries of conifers. We present techniques and related potential problems encountered during construction of a representative black spruce genomic library in the bacteriophage vector EMBL3. Molecular cloning manuals such as the following can be consulted for a detailed description of some of the cloning techniques presented here:

1) Maniatis et al. 1982, Molecular Cloning: a laboratory manual
2) Davis et al. 1980, Advanced bacterial genetics
3) Methods in Molecular Biology, Volume 2. 1984, and
4) Methods in Enzymology, Volume 152. 1987

2. ISOLATING HIGH MOLECULAR WEIGHT CONIFER DNA

Isolation of high molecular weight DNA from plant material is considerably more difficult than from animal tissues due in part to the presence of a rigid cell wall and high polysaccharide content (Miksche and Dhillon 1981). To be useful for library construction, three major criteria should be met by the nucleic acid isolated from plant material. First, the DNA must be pure enough to be cut by restriction endonucleases. Second, the yield of nucleic acid must be sufficient to produce a complete library with the available tissue. Third, the DNA must be of high molecular weight (at least 100 kb fragments) so restriction endonuclease cleavage will generate sufficient fragments of the desired size with cohesive ends.

A modification of the CTAB procedure (Murray and Thompson 1980) has been used successfully to isolate pure high molecular weight DNA from mature needles and seedling tissue of white spruce *Picea glauca*, black spruce *Picea mariana*, and jack pine *Pinus banksiana*.

2.1 DNA isolation using modified CTAB procedure

i) Cut 15-150 g of fresh needles. Wash needles in detergent, followed by surface sterilization in 2% hypochlorite solution for 20 minutes. Rinse three times with sterile distilled water.

ii) Freeze needles in liquid nitrogen and grind to a fine powder. Grind tissue in a large mortar, with pestle (freeze mortar slowly before adding liquid nitrogen to prevent breakage). About 10 g of tissue is added to the mortar, along with 10-20 mL liquid nitrogen. As soon as the tissue is frozen, it is ground for one minute to produce a powder. Needle fragments will be present; therefore, add 10 mL more of liquid nitrogen to the powder and regrind until no needle fragments remain. The powder must be free of needle fragments and finer than table salt.

iii) Add the powder to 100-200 mL cold 'CTAB extraction' buffer in a beaker. Mix gently.

iv) Repeat steps (ii) and (iii) until all needles are ground. Each batch of powder is added to the same beaker of CTAB extraction buffer so that the final volume of buffer plus powder is about 30 mL.

v) Filter solution through several layers of cheese cloth and one layer of miracloth. Re-extraction of the ground tissue increases DNA yield.

vi) Pellet organelles at 9 000 rpm for 15 minutes at 4°C. Discard supernatant.

vii) Resuspend pellet in a minimal volume (5-50 mL) 'CTAB wash' buffer. Bring to room temperature (22-25°C).

viii) Add 1/5 volume of 5% sarcosyl; keep 10 minutes at room temperature. The solution should become viscous. Then add 1/7 volume 5 M NaCl. Mix by gently inverting the tube several times. Add 1/10 volume 8.6% CTAB in 0.7 M NaCl. Mix gently.

ix) Incubate at 60°C for 15 minutes.

x) Add an equal volume of chloroform:isoamyl alcohol (24:1) and mix by inverting until emulsion is formed.

xi) Centrifuge at 5 000 rpm for 10 minutes; remove upper aqueous phase to clean tube.

xii) Add 2-3 volumes of cold absolute ethanol and mix by inverting quickly to precipitate the DNA. Hook out DNA using a bent Pasteur pipette and dissolve in a small volume of TE buffer. Allow DNA to dissolve at room temperature.

xiii) Add 0.96 g/mL CsCl and gently dissolve. Add ethidium bromide to a concentration of 300 µg/mL. Fill and seal appropriate sized ultracentrifuge tube. Spin at 40 000 rpm for a minimum of 16 hours at 15°C in an ultracentrifuge.

xiv) Remove DNA by puncturing tube 2 mm below the band with an 18 gauge needle and dripping DNA band into a sterile tube.

xv) Add salt-saturated isoamyl alcohol, or water saturated isobutanol, and extract the collected band until no ethidium bromide remains in the lower aqueous DNA layer. Extract three more times.

xvi) Dialyze against at least three changes of 4 L each of TE buffer at 4°C. If necessary concentrate DNA sample by successive gentle extractions with an equal volume of isobutanol. Redialyze, or precipitate the DNA by adding 1/10 volume 3M sodium acetate and 2-3 volumes ice cold

absolute ethanol. Centrifuge at 12 000 rpm to pellet the DNA, discard supernatant, and wash pellet with 70% cold ethanol. Centrifuge, discard supernatant, dry pellet in a vacuum dessicator, and resuspend in TE buffer.

xvii) Determine the DNA concentration by optical density reading at wavelength 260 nm or gel electrophoresis against DNA standards. Yield for black spruce is about 900 µg/100 g needles. Yield for jack pine is about 250 µg/100 g needles.

xviii) Analyze the DNA by electrophoresis at room temperature in a 0.4% agarose NA (Pharmacia) gel with intact λ size marker. High molecular weight DNA will migrate as a tight band behind the λ DNA.

2.1.1 Precautions:

1) Tissue must be frozen as quickly as possible and thawed only in the presence of nuclease inactivating agents such as detergent, EDTA, or phenol.

2) Repeated freezing of high molecular weight DNA should be avoided. Sterile DNA solutions containing no active enzymes can be stored for long periods at 4°C in TE buffer (Kaiser and Murray 1985).

3) Mechanical shearing of DNA must be kept to a minimum. Broken fragments inhibit various stages of the cloning procedure by joining nonproductively to vector DNA fragments, interfering with ligation, and inhibiting packaging reactions. Use wide mouth pipettes when handling high molecular weight DNA solutions.

4) Plant DNA isolated by this procedure has ratios of uv absorbance at 260/280 nm of about 1.8-1.9 and 260/230 nm of about 2.0-2.4. High protein or polysaccharide content of the starting material may result in nucleic acid contaminated with these substances. These contaminants may lead to problems with resuspending ethanol precipitated DNA and/or digesting DNA with restriction endonucleases.

2.1.2 *Troubleshooting*

Problem	Possible Causes	Solution
low DNA yield	a) cellular organelles not released	a) grind tissue finer
	b) nuclease activity	b) i) keep tissue frozen during extraction and immediately transfer ground tissue to cold extraction buffer ii) sterilize tissue prior to grinding iii) sterilize all solutions
	c) age/type of tissue	c) sources of tissue for higher yields are cell suspensions, seedlings, and 1-year-old needles from mature trees
DNA won't dissolve	a) protein content too high	a) i) re-extract with chloroform:isoamyl alcohol (24:1) ii) increase volume of 'CTAB wash' buffer iii) reband DNA on CsCl gradient
	b) DNA concentration too high	b) add more TE buffer
	c) solution needs heat or time to dissolve	c) i) heat solution for 10 minutes or less at 65°C ii) incubate solution at 4°C and allow to dissolve for 1-2 days
DNA degraded	a) possible exonuclease or microbial contaminination	a) treat DNA with 0.1% SDS and 1/100 volume protease K (3 mg/mL) at 37°C for 30 minutes. Follow with phenol/chloroform extraction and ethanol precipitation
	b) mechanical shearing occurred during isolation procedure	B) i) avoid freeze-thaw cycles ii) use wide-bore pipette tips iii) handle DNA gently during resuspension

Note: The modified CTAB method can be used to isolate DNA from seedling tissue. Less starting material is required to obtain a DNA yield equivalent to the yield from mature needle tissue.

2.2 Genomic DNA isolation from tissue cultured plant cells

Tissue cultured plant cells are an excellent source of DNA for cloning because of the relative ease of extraction, compared with grinding mature needles, and because of the higher yield of DNA per gram of starting material. Because the tissue is sterile there are no fungal or bacterial enzymes, and the cells are lysed readily without major preliminary extractions. The yield per gram of tissue cultured cell material can be as much as tenfold (our unpublished data: 900 µg DNA/100 g needle tissue contrasted to 1.4 mg DNA/10 g tissue cultured cells) above that of mature tissues such as conifer needles. The following method, adapted from Mettler 1987, has been used successfully in our laboratory to isolate DNA from embryogenic cells and non-embryogenic callus tissue of various conifer species.

DNA isolation procedure for tissue cultured cells

i) Weigh tissue, usually 10 g, and transfer to a sterile mortar. If using less than 100 mg of tissue transfer instead to a small ground glass homogenizer to reduce loss of homogenate on vessel surface. Add two volumes Mettler Homogenization buffer and homogenize without causing excessive foaming. Transfer homogenate to centrifuge tube and rinse homogenizing vessel with 1 to 2 volumes of Mettler Homogenization buffer.

ii) Incubate homogenate under sterile conditions at room temperature for an additional 20 to 30 minutes, grinding occasionally to break any remaining unlysed cells. Add an equal volume of phenol (equilibrated with 0.5 M NaCl and 100 mM Tris.HCl, pH 8.0); mix until an emulsion is formed and centrifuge to separate layers. Carefully remove the upper aqueous phase and mix until emulsion is formed with phenol:chloroform:isoamyl alcohol (25:24:1). Centrifuge to separate layers.

iii) Collect upper aqueous phase and precipitate DNA with 1/10 volume 3 M sodium acetate and 2 volumes of cold absolute ethanol. Collect DNA by hooking out precipitate or pelleting by centrifugation. Resuspend DNA in TE buffer, then band on CsCl equilibrium gradient (refer to steps (xii) to (xvii) of modified CTAB procedure, Section 2.1).

iv) Yield is about 1.4 mg DNA/10 g frozen cells.

3. FRACTIONATION OF DONOR DNA

The substitution vector EMBL3 will accept inserts of foreign DNA ranging from 9-23 kb in size (Frischauf et al. 1983). To generate a population of molecules with an average size of 20 kb for cloning into this kind of vector, eucaryotic DNA can be prepared by either (1) random fragmentation, or (2) restriction endonuclease digestion under controlled conditions. A truly random library can only be obtained from DNA that has been fractionated in a sequence-independent manner, such as by mechanical shearing. However, this method requires further enzymatic manipulations to produce cohesive ends compatible with the vector arms. It is more convenient to fractionate DNA by partial digestion with restriction enzymes because this will produce random overlapping fragment sets sufficient for most purposes. A library generated with these fragments is not truly random because (a) the termini of the fragments are not distributed randomly among the potential cleavage sites, and (b) these sites are not distributed randomly along the length of the genomic DNA.

Fractionation of genomic DNA by partial digestion is usually accomplished with the restriction endonuclease *Mbo*I or its isoschizomer *Sau*3A. These enzymes produce cohesive ends compatible with the EMBL3 vector arm termini that are generated when the central stuffer fragment is excised using the enzyme *Bam*HI. Genomic DNA fragments in the desired size range can be separated physically from other fragments by various methods. Gel electrophoresis with DNA elution provides suitable DNA, although there is potential inhibition of ligase activity from agarose or other contaminants. A preferable alternative is velocity centrifugation through sucrose or sodium chloride gradients.

3.1 Preparation of 20 kb fragments of eucaryotic DNA

All enzymes should be checked for the presence of appropriate activity and be free of contaminating nucleases that may interfere with subsequent ligation of DNA fragments. Use only enzymes that are supplied with test results for endonuclease assay, exonuclease assay, and ligation/recut assay (e.g. those supplied by BRL Life Technologies Inc., Gaithersburg, Md.

3.1.1 Fractionation of donor DNA

i) Mix 10 µg of genomic DNA with 10 µL of 10X *Mbo*I (or *Sau*3A) buffer and add sterile distilled water to make a total volume of 100 µL.

ii) Dispense 20 µL into a microcentrifuge tube (tube 1) and 10 µL into tubes 2-9. Cool tubes on ice.

iii) Adjust commercial enzyme preparation of *Mbo*I (or *Sau*3A) to a concentration of 1 unit/µL using manufacturer's suggested buffer. Add 1 µL to tube 1 and mix (0.5 units per µg). Remove 10 µL and mix with contents of tube 2 (0.25 units per µg). Continue serial dilution until tube 8, mixing thoroughly at each step. Tube 9 is an undigested DNA control.

iv) Incubate at 37°C for one hour.

v) Stop reaction by immediately placing samples on ice and adding EDTA to a final concentration of 20 mM. Samples are run on a 0.4% TAE agarose gel and separated by horizontal electrophoresis at 2 V per cm. Run λ DNA cut with *Eco*RI or *Hind*III as a size marker. Low voltages improve resolution of high molecular weight DNA.

vi) Determine the concentration of enzyme that produces maximum fluorescence in the 18-23 kb size range. Half that concentration of enzyme will produce the maximum number of molecules in the desired size range.

vii) Scale up the reaction to digest required amount of DNA to produce a representative library. Maintain all variables as in the trial reaction. Choose enzyme concentrations equal to half of and twice the optimum to fractionate genomic DNA, so as to minimize misrepresentation of fragments containing a non-random distribution of cleavage sites.

viii) Stop the reaction immediately by placing the reactions on ice and subsequently adding EDTA to a concentration of 20 mM. Analyze aliquots of cleaved DNA by electrophoresis in agarose gel with appropriate size markers and undigested DNA, to check size distribution.

ix) Pool samples, gently extract twice with phenol:chloroform: isoamyl alcohol (25:24:1). Add 1/10 volume 3 M sodium

acetate and precipitate the DNA by adding 2 volumes cold absolute ethanol. Pellet DNA by centrifugation, wash pellet twice with 70% ethanol, and resuspend in 200-500 µL TE buffer.

x) Purify sized fragments to remove any compounds that may inhibit ligation and/or packaging reactions (see Appendix I).

3.1.2 *Troubleshooting*

Problem	Possible Causes	Solution
DNA won't cleave	a) inhibitors or impurities present in DNA solution	a) i) extract solution twice with phenol and twice with chloroform:isoamyl alcohol (24:1). Precipitate DNA with 1/10 volume 3 M sodium acetate and 2 volumes ethanol. Resuspend precipitate in TE buffer and repeat partial digestion. ii) reband DNA on CsCl gradient
	b) DNA not evenly resuspended in solution when trial was done	b) heat DNA to 65°C for 10 minutes. Check that the DNA is evenly resuspended by carefully pipetting solution.
DNA degraded	a) exonuclease or microbial contamination present in DNA solution	a) treat DNA with 0.1% SDS and 1 drop of protease K (3 mg/mL) at 37°C for 30 minutes. Follow with phenol/chloroform extraction and ethanol precipitation
	b) addition of too much enzyme	b) see b) above

4. VECTOR PREPARATION

The bacteriophage λ replacement vector EMBL3 has a cloning capacity of 9-23 kb. It is the product resulting from the insertion of multiple cloning sites between the essential lambda arms and non-essential central polystuffer recognition sites of λ 1059 (Karn et al. 1980). EMBL3 multiple cloning sites contain recognition sequences for the restriction enzymes *Bam*HI, *Sal*I, and *Eco*RI (Figure 2). Flanking these sites are the vector right and left arms. Virion proteins are encoded on the left arm which is 19.3 kb in size. The right arm encodes the origin of replication, major promoters, and a number of other essential genes. At the ends of the arms are 12-base complementary single-stranded projections. These cohesive (cos) sites permit the λ genome to circularize following infection and are regenerated when the phage genomes are packaged. Between the arms is the central fragment of EMBL3 that is flanked by polylinkers in inverse orientation, as follows:

*Sal*I-*Bam*HI-*Eco*RI--central stuffer region--*Eco*RI-*Bam*HI-*Sal*I

Any of the three enzymes (*Sal*I, *Bam*HI, or *Eco*RI) can be used for cloning. However, the vector was designed (Karn et al. 1980) especially for use with *Bam*HI which makes a staggered cut at the hexanucleotide sequence:

```
    ↓
G G A T C C
C C T A G G
        ↑
```

The core of the *Bam*HI recognition sequence is the recognition site of the enzymes *Mbo*I and *Sau*3A, which cleave

```
↓
G A T C
C T A G
      ↑
```

Thus, EMBL3 DNA cleaved by *Bam*HI can accept an exogenous DNA fragment cleaved by partial digestion with *Mbo*I or *Sau*3A.

A complete eucaryotic genomic library generally consists of millions of infective phage particles, each of which has a unique fragment of the eucaryotic DNA. Only EMBL3 phage genomes in the 40-52 kb size range are packaged efficiently into infective phage particles. Ligation of the two arms of a replacement vector generates a DNA molecule that is too small to be packaged. Therefore, theoretically, only molecules that have an exogenous DNA fragment of 17-23 kb joined to a vector left and right arm (recombinant DNA molecule) will be packaged and infectious. This provides an enrichment technique for recombinant molecules. Also, the EMBL3 vector has certain genes whose phenotype can be utilized to select for recombinant phage clones. Selection is based on the Spi^+ (susceptible to P2 inhibition) phenotype. Growth of phage displaying this phenotype is restricted on bacterial strains that are lysogenic for prophage P2. The genes responsible for P2 inhibition are present on the central stuffer region of EMBL3 vectors. When these genes (*red* and *gam*) are removed and replaced with insert DNA the phage expresses the Spi^- phenotype and can be plated on a host containing the P2 lysogen.

4.1 Purification of intact phage

Single plate stocks are grown from individual plaques and titered by serial dilution plating. Large-scale phage isolation by double banding in CsCl equilibrium gradients gives the purest phage preparations. The following method, adapted from Davis et al. 1980, uses this technique.

Isolation of phage

i) To purify the phage, streak out the phage on a fresh LB plate as described by Davis et al. 1980, using *E. coli* host strain lacking McrA and McrB genetic loci (see section 5 for explanation).

ii) Remove plugs of several individual turbid plaques, grow individual plate stocks from them using fresh LB plates, and titer.

iii) Select that stock as the starter stock and prepare ten 150 mm plate stocks, as described by Davis et al. 1980. Use 2.5 X 10^6 plaque-forming units (pfu), 125-150 µL host plating cells, and 5 mL LB soft agar per 150 mm petri dish. After 5-6 hours of incubation at 37°C the plates should have a mottled-looking lawn. If left to grow longer than 5-6 hours the titer of the plate stock will drop. If phage is plated on NUNC plates (530 cm^2) use

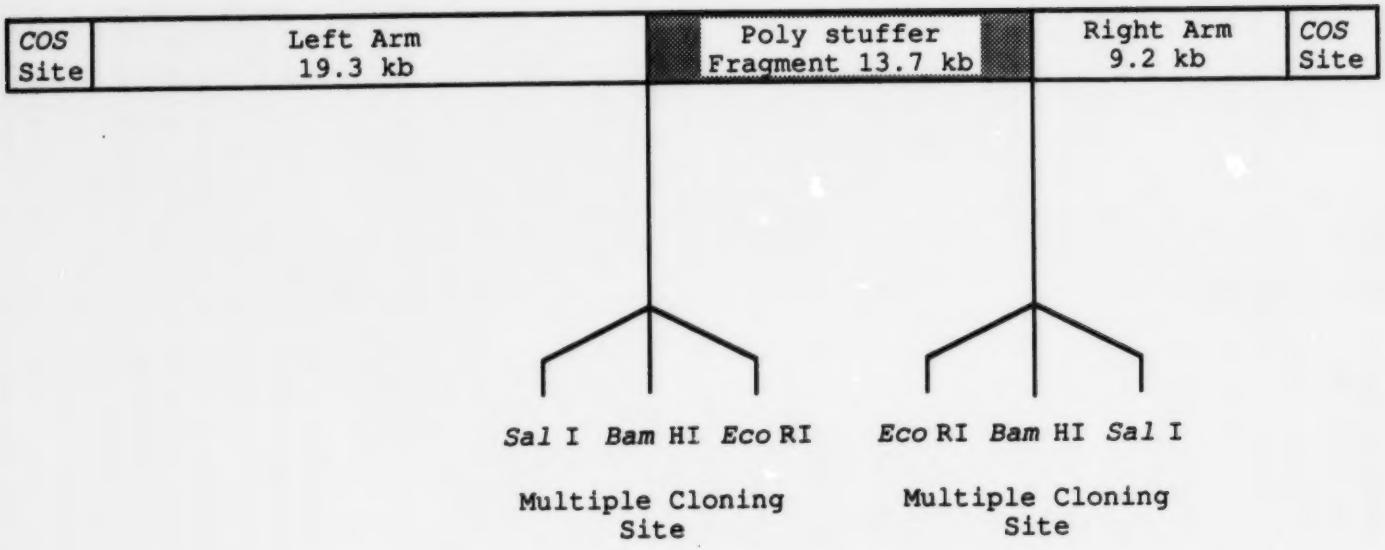


Figure 2. Simplified map of the EMBL3 vector

7.5×10^6 pfu/plate, 1 mL host plating cells, and 35 mL top agar.

iv) Cool plates at 4°C at the point when the lawn looks the lightest. When plates have cooled thoroughly, overlay with 10 mL cold λ diluent plus a few drops of chloroform. Let phage diffuse into λ diluent at 4°C overnight.

v) Remove overlay solution and sediment debris by centrifugation in an ultracentrifuge at 10 000 rpm at 4°C for 15 minutes.

vi) Pellet phage by ultracentrifugation at 23 000 rpm at 4°C for 90 minutes.

vii) Resuspend each phage pellet overnight in 1 mL λ diluent.

viii) Add 0.75 g CsCl/mL of phage suspension. Centrifuge phage for a minimum of 5 hours at 54 000 rpm in an ultracentrifuge. Remove phage band from gradient and reband at 54 000 rpm. Collect the phage band. Dialyse phage solution against TE buffer to remove CsCl. Phage can be stored at 4°C.

4.2 Extraction and purification of phage DNA

i) To 500 µL phage suspension add 10 µL 10% SDS plus 5 µL protease K (3 mg/mL in 50% glycerol). Incubate at 37°C for at least 30 minutes.

ii) Extract twice with phenol:chloroform (1:1) and once with chloroform:isoamyl alcohol (24:1) followed by ethanol precipitation. Resuspend in 500 µL TE. Note that the DNA usually is concentrated and needs to resuspend passively at 4°C for 2-3 days to become a homogeneous suspension.

4.3 Fractionation of vector DNA and purification of phage arms

Bacteriophage λ vectors, such as EMBL3, used in library construction require that the central polystuffer segment of their genomes be removed to accommodate 15-20 kb of foreign DNA. In EMBL3 the *Bam*HI sites are flanked internally by *Eco*RI sites. Cleavage with both *Bam*HI and *Eco*RI releases short (<10 bp) *Bam*HI-*Eco*RI linker fragments from both ends of the central fragment. This prevents competition between the central fragment and donor DNA for ligation sites on the vector arms. As a general rule for EMBL3 vector DNA, first digest with *Bam*HI until cleavage is complete and then adjust the reaction buffer to *Eco*RI conditions and digest with a predetermined amount of *Eco*RI.

4.3.1 *Cleavage of vector DNA*

i) Determine the conditions for complete digestion of vector DNA with *Bam*HI and *Eco*RI in separate small scale pilot reactions. Analyze the cleaved DNA for complete digestion by agarose gel electrophoresis.

ii) Digest 20-50 µg of vector DNA with the predetermined amount of *Bam*HI and check that the reaction is complete by electrophoresis of a small sample (0.2 µg or less).

iii) When *Bam*HI digestion is complete adjust the reaction buffer to *Eco*RI conditions and digest with a predetermined amount of *Eco*RI.

iv) Extract gently once with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform. Add 1/10 volume 3 M sodium acetate and precipitate the DNA with isopropanol to leave the 10 bp linker in solution.

v) Collect the precipitate by centrifugation, wash twice with 1/2 volume 40% isopropanol plus 1/10 volume 3 M sodium acetate, dry briefly in a vacuum dessicator, and resuspend in TE buffer at an approximate concentration of 250-500 µg/mL. Determine exact DNA concentration and purify by NACS column purification if required (see Appendix I).

4.3.2 *Treatment of fractionated vector DNA with alkaline phosphatase (optional) (Maniatis et al. 1982)*

Calf intestinal alkaline phosphatase catalyzes the removal of 5′ phosphate residues from DNA and RNA. Removing the 5′ phosphates from the vector fragments prevents religation of the polystuffer and linker fragments to vector arms. This, in turn, reduces competition between these fragments and the eucaryotic DNA insert for the vector arms.

i) To 45 µL fractionated vector in TE buffer add 5 µL 10X CIP buffer and 1 µL CIP

(calf intestinal alkaline phosphatase, 2.2 units/µL).

ii) Incubate at 37°C for 30 minutes, add a second aliquot of CIP, and continue incubation for a further 30 minutes.

iii) Add 40 µL of H_2O, 10 µL of 10X STE, and 5 µL of 10% SDS. Heat to 68°C for 15 minutes.

iv) Extract twice with phenol:chloroform (1:1) and twice with chloroform:isoamyl alcohol (24:1).

v) Precipitate DNA by adding 1/10 volume 3 M sodium acetate and 2 volumes cold absolute ethanol.

vi) Centrifuge at 12 000 rpm for 10 minutes minimum to pellet DNA and discard the supernatant. Wash DNA pellet with 70% ethanol, then centrifuge and resuspend pellet in TE buffer to a final concentration of 0.5 µg/µL.

4.3.3 *Troubleshooting*

Problem	Possible Causes	Solution
vector DNA won't cleave	a) inhibitors/impurities in DNA solution	a) i) re-extract solution with phenol followed by chloroform:isoamyl alcohol (24:1) extraction. Precipitate DNA with 1/10 volume 3 M sodium acetate and 2 volumes ethanol. Resuspend in TE buffer and recleave ii) precipitate the DNA with spermidine (refer to Appendix I-D)
	b) Insufficient enzyme used	b) run small pilot reactions to determine necessary number of enzyme units required
vector DNA is degraded	a) solutions contaminated	a) i) prepare fresh buffers and distilled water, then autoclave ii) work under sterile conditions
	b) vector DNA is contaminated with exonucleases	b) obtain purified, exonuclease-free vector DNA

5. Significance of host cell strain when cloning plant genomic DNA

Plant genomic DNA contains significant amounts of 5-methycytosine. Over 25% of the cytosine (C) residues are methylated (Shapiro 1976) in plants, whereas 5-methylcytosine accounts for about only 2-7% of the total cytosine in animal cells (Razin and Riggs 1980). In plants, methylation occurs at cytosine-guanine (C-G), cytosine-thyamine (C-T), cytosine-cytosine (C-C), and cytosine-adenine (C-A) sites. In addition, the trinucleotide sequences C-A-G and C-T-G contain significant levels of methylated cytosine as shown in Appendix IV (Greunbaum et al. 1981).

To construct a representative library of plant genomic DNA all the DNA fragments must clone with similar efficiency. However, several *E. coli* strains commonly used as hosts for construction of genomic libraries in bacteriophage vectors contain genes that code for enzymes which restrict methylated DNA via double-stranded cleavage. The result of restriction is death of the phage carrying the methylated DNA. The genetic loci associated with these enzymes are designated McrA, McrB, and Mrr (Raleigh et al. 1988).

Mcr (for modified cytosine restriction) genetic loci in the *E. coli* host affect cloning of highly methylated plant genomic DNA as follows:

1) Strains of *E. coli* with Mcr loci are not transfected efficiently because recombinants containing methylated sequences are restricted. To construct a plant genomic library, for example of black spruce, the number of recombinants required to represent the genome is 3 million. Therefore, inefficient transfection is costly.
2) There will be an underrepresentation of those genomic DNA sequences that contain methylated cytosine.
3) Chloroplast and mitochondrial DNA in many plants is unmethylated (Tewari and Wildman 1966, Bonen et al. 1980); therefore, there will be an overrepresentation of recombinants containing chloroplast-derived and mitochondrial-derived inserts in plant libraries plated on commonly used *E. coli* strains.

To avoid misrepresentation of sequences with methylated cytosine the *E. coli* host strain for library construction should lack functional Mcr genetic loci. A detailed list of the Mcr phenotypes of a number of *E. coli* strains appears in the paper by E.A. Raleigh et al. 1988.

E. coli strain WA802 ($McrA^-$ and $McrB^-$) was used in our laboratory to construct a genomic library of *Pinus banksiana* consisting of eight million recombinants. This strain was a gift from D.A. Johnson, University of Ottawa, Department of Biology.

6. LIGATION, PACKAGING, AND TITERING

It is necessary to determine the ratio of arms to potential inserts when setting up a ligation reaction. Optimal values for each parameter can be estimated theoretically (Maniatis et al. 1982). However, a proportion of the molecules will have damaged ends, so actual values rarely match theoretical ones. Therefore, pilot reactions are carried out in a series of test ligations to check ligation efficiency. The simplest approach is to analyze each variable independently in a series of test ligations.

Ligation of the cohesive ends of EMBL3 digested with *Bam*HI is carried out at temperatures between 10°C and 16°C. Hydrogen bonds between the cohesive ends of the DNA fragments to be ligated are more stable at lower temperatures (Dugaiczyk et al. 1975). The concentration of ATP in the ligation reaction is also critical. Ligation of blunt ends is completely inhibited at a final ATP concentration of 5mM. However, cohesive end ligation is only inhibited by 180 mM final ATP concentration (Ferretti and Sgaramelle 1981). The optimal ATP concentration for both blunt end and cohesive end ligation is between 0.5 and 1mM.

6.1.1 Determination of optimal ligation conditions

i) At all stages of testing maintain constant all variables such as ligase concentration, buffers, packaging extracts, and host strains. Pre-anneal cohesive ends by incubation of cleaved vector DNA at 42°C for one hour in the presence of 10 mM $MgCl_2$ before ligation.

ii) Ligate varying amounts of vector DNA to a constant amount of donor DNA (0.5 µg) at weight to weight ratios of 1:1, 2:1, 3:1. Maximum efficiency should be obtained with roughly equal molar ratios of annealed vector arms and donor DNA fragments. Vector arms are 30 kb and donor DNA is at average 20 kb, so equal molar ratio is 3:2.

iii) Divide packaging mix into five equal volumes and add, to each, 1/5th volume of the total ligation reaction mixture. Package according to supplier's instructions and titer.

iv) The volume of the ligation reaction is kept to 5-10 µL for best efficiency of both ligation and packaging. Note that efficiency of packaging is improved if the ligated DNA is phenol extracted, ethanol precipitated, and resuspended in 5-10 µL of TE buffer prior to packaging in 50 µL packaging extract.

6.1.2 Troubleshooting

Problem	Possible Causes	Solution
donor DNA won't ligate to vector DNA	a) donor DNA solution contains inhibitors	a) purify DNA (see Appendix I)
	b) ligase buffer is incorrect	b) replace buffer
	c) T4 ligase inactive	c) replace ligase
	d) either vector or donor DNA (or both) is degraded	d) electrophorese DNA in agarose gel to determine if DNA is degraded. If so, re-isolate, purify, and proceed
DNA not packaging	a) vector DNA not annealed	a) prior to ligation anneal vector DNA by incubating at 42°C for one hour in the presence of 10 mM $MgCl_2$
	b) inhibitors present	b) i) purify DNA (see Appendix I) ii) isolate DNA from alternate tissue or developmental stage if possible
	c) ligation didn't work	c) check ligations by electrophoresis, then refer to above section on why donor DNA won't ligate to vector DNA
	d) controls indicate packaging malfunction	d) replace batch or try a new brand
low titers	a) ligation or packaging didn't work	a) see above
	b) phage didn't adsorb to cells	b) grow cells in presence of 0.2% maltose and 10 mM $MgSO_4$
	c) host bacterial cells not competent	c) i) check medium composition for maltose ii) pellet overnight culture of host cells and resuspend in 10 mM $MgSO_4$ iii) check growth conditions; temperature, volume of medium, size of aeration container iv) grow fresh streak of host from -80°C glycerol stock. Pick single colony for overnight broth culture. v) check cells with phage to determine if cells are true to type
	d) ligation ratio not optimal	i) try alternate ratios of vector to donor DNA such as 1:2, 1:3, 1:4 if DNA quantity is not limiting.
tiny plaque morphology	a) top agar concentration too high	a) prepare top agar at concentration of 0.6% or less
	b) Mg^{++} concentration too low/high	b) adjust $MgSO_4$ to final concentration of 10mM in TB plating and top agar

6.2 Large scale ligation and packaging

Once the optimal ratio of vector to donor DNA has been determined and background taken into account, multiple large scale ligation reactions are set up using the same parameters as determined by the test ligations. It is important to maintain constant ratios of all solutions in larger scale. From titers obtained in the test ligations it is possible to estimate the quantities of both vector and donor DNA required to produce a complete library.

i) Grow an overnight culture of both permissive (NM538 or LE392) and restrictive (NM539) hosts by inoculating a single colony into 50 mL of TB broth in a 250 mL flask supplemented with 0.5 mL of 20% maltose and 0.5 mL 1 M $MgSO_4$. Shake overnight at 30°C. Next day pellet the cells and resuspend in 2/5 volume 10 mM $MgSO_4$. Store at 4°C until required.

ii) Ligation reaction:

* x µg EMBL3 DNA in TE buffer
* y µg donor DNA
10.0 µL 5X ligase buffer freshly mixed (see Appendix II)
10.0 µL T4 DNA ligase (10 Weiss units)
water to 50 µL

* x + y = maximum of 5 µg

This reaction can be scaled up but volume must be kept equal to or below 50 µL or efficiency of reaction will be compromised. Set up as many reactions as required to produce the complete library.

iii) Incubate ligation at 15°C overnight.

iv) Package 10 µL of the ligation (maximum 5 µg EMBL3 DNA) in one packaging extract. Package ligated DNA according to the protocol supplied with the in vitro packaging system used. Efficiencies of packaging can be increased by precipitating ligated DNA with 1/10 volume 3 M sodium acetate and cold absolute ethanol and resuspending the DNA in TE buffer prior to packaging.

7. AMPLIFICATION AND STORAGE

Amplified libraries will maintain a high titer for several years. They are useful as a source from which many genes may be isolated over a period of time, and also for repeated use such as chromosome walking. Amplified libraries are convenient if a library is to be distributed to other researchers. One disadvantage of amplification is that differential growth of recombinants will be reflected in the sequence content of the amplified library and, consequently, the number of clones necessary for screening to find specific sequences is increased. Also, proportional misrepresentation occurs because of competitive growth of recombinants in liquid culture or at high plaque density.

Misrepresentation of sequences in the amplified library happens when the host bacterium (e.g. NM539 rec+) is simultaneously infected with more than one phage clone, allowing recombination events to occur. Plating the primary library at relatively low density as plaques on agar plates, such that individual plaques can still be resolved, minimizes the problem. To reduce heterogeneity of plaque size and hence differential amplification, use fresh plating cells to which phage have been preadsorbed and keep plates horizontal until top agar has set. Prevent condensation from smearing the plaques by inverting plates during incubation or drying plates under sterile conditions prior to incubation.

Amplification and storage protocol

i) Titer all packaged phage extracts as accurately as possible.
ii) On a 530 cm^2 NUNC culture dish plate phage to a concentration of 50-100 X 10^3 pfus/plate.
iii) Incubate and allow plaques to develop for 7-12 hours until distinct plaques are formed. The plaques should be well separated to prevent recombination events.
iv) Incubate plates at 4°C for 1 hour, add 35 mL SM buffer, and incubate at 4°C overnight in horizontal position.
v) Collect buffer, add chloroform, shake gently, and pellet debris.
vi) Remove lysate to fresh tube and store at 4°C over chloroform indefinitely.

8. EVALUATION OF GENOMIC DNA LIBRARY QUALITY

The ability to productively isolate a cloned gene from a genomic DNA library is one of the most important determinants of the quality of that library. Subsequent to amplification the library should be evaluated to determine the abundance of sequences in the library DNA, and whether major rearrangements have occurred during library construction and amplification. An empirical method, descibed by Phillips et al. 1985, is available to determine whether a λ genomic library will be a productive source of clones carrying specific sequences. Comparative analysis by blot hybridization demonstrates that a library will be a productive source of clones when corresponding bands are present in library and genomic DNA digested with the same restriction enzyme. Absence of a band in the library sample suggests that the sequence of interest is not present in the library. Presence of an unique band in the library that does not correspond to a band in the genomic DNA digest suggests the possibility of genomic DNA rearrangements during library constuction. Details of this evaluation procedure can be found in Phillips et al. 1985.

9. CONCLUDING REMARKS

Construction of a genomic DNA library provides a pool of genes and their related sequences which can be used to study nucleotide structure, regulation and expression of genes. In this manual we have described the construction of a black spruce bacteriophage library in the EMBL3 vector. DNA from coniferous tree species is more difficult to isolate, purify, and clone than DNA from sources such as bacteria or animal tissue. This may be due in part to factors such as the protein and polysaccharide content of the starting material (needles or seedlings), and the degree of methylation or stereochemistry of the DNA molecule. A genomic DNA library of lodgepole pine has been constructed from seedling tissue in the EMBL3 vector using Packagene (Promega) extracts (unpublished results; Byrne et al.). However, packaging in Packagene extracts was inhibited when EMBL3 was ligated to genomic black spruce DNA isolated from mature needle or seedling tissue. To overcome this problem EMBL3 was ligated to black spruce genomic DNA isolated from cell suspension cultures and packaged in GigaPak Gold extracts (Bio-Mega).

Highly conserved ubiquitous genes previously isolated from other species can be used to 'probe' libraries for related sequences. Comparison of nucleotide sequence data of genes from different species can potentially be used to understand the molecular biology of different species, evolutionary mechanisms, and relationships among species. A permanent library of clones bearing overlapping sequences will facilitate isolation of many linked genes that constitute a complex genetic locus and those that are coordinately and/or sequentially expressed. Further investigations should reveal possible interrelationships between cellular events and gene expression. By studying effects of external factors such as disease, nutritional status, environmental stress, and developmental changes on gene expression insight can be gained into the basic molecular mechanisms in trees.

References

Berger, S.L.; Kimmel, A.R. 1987. Guide to molecular cloning techniques. Methods in Enzymology Vol. 152. Academic Press, Inc.

Bonen, L.; Huh, T.Y.; Gray, M.W. 1980. Can partial methylation explain the complex fragment patterns observed when plant mitochondrial DNA is cleaved with restriction endonucleases. FEBS Letters 111: 340-346.

Blumenthal, R.M. 1986. *E. coli* can restrict methylated DNA and may skew genomic libraires. Trends Biotechnol. 4: 302-305.

Clarke, L.; Carbon, J. 1976. A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. Cell 9: 91-99.

Davis, R.W.; Botstein, D.; Roth, J.R. 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Box 100. Cold Spring Harbor, NY.

Dugaiczyk, A.; Boyer, H.W.; Goodman, H.M. 1975. Ligation of *Eco*RI endonuclease generated DNA fragments into linear and circular structures. J. Mol. Biol. 96: 171-184.

Ferretti, L.; Sgaramelle, V. 1981. Specific and reversible inhibition of the blunt end joining activity of the T4 DNA ligase. Nucleic Acids Res. 9: 3695-3705.

Frischauf, A-M.; Lehrach, H.; Poustka, A.; Murray, N. 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170: 827-842.

Gruenbaum, Y.; Naveh-Many, T.; Cedar, H.; Razin, A. 1981. Sequence specificity of methylation of higher plant DNA. Nature 292: 860-862.

Kaiser, K.; Murray, N.E. 1985. The use of phage λ replacement vectors in the construction of representative genomic DNA libraries. Pages 1-47 *in* Glover, D.M., ed. DNA cloning: a practical approach. IRL Press, Oxford.

Karn, J.M.; Brenner, S.; Barnett, L.; Cesareni, G. 1980. Novel bacteriophage lambda cloning vector. Proc. Natl. Acad. Sci. (USA) 77: 5172-5176.

Maniatis, T.; Fritsch, E.F.; Sambrook, J. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, NY.

Mettler, I.J. 1987. A simple and rapid method for minipreparation of DNA from tissue cultured plant cells. Plant Mol. Biol. Reporter 5: 346-349.

Miksche, J.P.; Dhillon. 1981. Nucleic acid research techniques in forest genetics. Pages 329-371 *in* Advances in Forest Genetics (India).

Murray, M.G.; Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8: 4321-4325.

Phillips, S.L.; Cassavant, N.C.; Hutchison, C.A., III; Marshall, H.E. 1985. An empirical method for the evaluation of the quality of genomic DNA libraries. Nucleic Acids Res. 13: 2699-2708.

Raleigh, E.A.; Murray, N.E.; Revel, H.; Blumenthal, R.M.; Westaway, D.; Reith, A.D.; Rigby, P.W.J.; Elhai, J.; Hanahan, D. 1988. McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. Nucleic Acids Res. 16: 1563-1575.

Razin, A.; Riggs, A.D. 1980. DNA methylation and gene function. Science 210: 604-610.

Rogers, S.O.; Bendich, A.J. 1985. Extraction of DNA from milligram amounts of fresh herbarium, and mummified plant tissues. Plant Mol. Biol. 5: 69-76.

Shapiro, H.S. 1976. CRC Handbook: Biochemistry and Molecular Biology. 2: 259.

Tewari, K.K.; Wildman, S.G. 1966. Chloroplast DNA from tobacco leaves. Science 153: 1269-1271.

Walker, J.M., ed. 1984. Methods in Molecular Biology 2: Nucleic Acids. Humana Press, Clifton, NJ.

Acknowledgements

The authors would like to thank F.M. Tremblay, P. Périnet, and J. Lavereau for providing conifer cell suspensions, D.A. Johnson for kindly providing strain WA802, and E.A. Raleigh for kindly providing strains ER1647 and ER1648. The authors are also grateful to V.L. Seligy and L. Bonen for reviewing this manual.

Appendix I: Methods for DNA Purification

A. ELUTIP-d method (from Schleicher & Schuell general instructions for use)

1. To elute 20 kb fragments from agarose gel, cut out the band of interest. Put band into a 20 mL syringe.
2. Force the agarose through an 18 gauge needle (with syringe plunger) into a sterile screw-cap tube. Add 10 mL Low Salt elution buffer (or 3 times agarose volume). Vortex.
3. Shake at 37°C overnight. The DNA is now in the buffer.
4. Place a GF/C filter at the bottom of a 20 mL syringe. Pour buffer with gel pieces into it. Allow buffer to flow through filter without additional force. Collect buffer in a clean tube.
5. Use syringe plunger to gently remove remaining liquid from the agarose in the syringe, collecting buffer in a microcentrifuge tube. Add this buffer to that collected in step 4.

To prepare the column

1. Load a 5 mL syringe with 1-2 mL High Salt solution.
2. Cut off tip of ELUTIP-d column. Remove top protector cap and attach to syringe.
3. Prewash the matrix by forcing High Salt solution into the column. Caution: ELUTIP-d column is designed for solution flow in one direction only.
4. To prime the column load a second syringe with 5 mL Low Salt solution.
5. Remove column from the first syringe and reattach to second syringe with Low Salt solution. Prime column with Low Salt solution to ensure that all the High Salt solution is washed from the column.

Binding and elution of DNA

1. After disconnecting column from the syringe load syringe with DNA sample.
2. Attach 0.45 µm cellulose acetate disposable prefilter to syringe. Use of prefilter insures removal of particulates from sample.
3. Reattach column to disposable filter holder, then pass all of the sample through the filter and column. DNA adsorbs to the matrix in the column. Slow flow is necessary for complete adsorption of the DNA to the matrix. Flow rate should be 1-2 mL/minute.
4. Disconnect column and filter from syringe (do not discard) and load sample syringe with 2-3 mL Low Salt solution. Reconnect column and filter to syringe and wash any remaining DNA sample from the syringe and filter.
5. Remove filter and column from syringe and discard prefilter.
6. To elute DNA load syringe with 0.4 mL High Salt solution and reattach column to syringe. Elute DNA with High Salt solution into a 1.5 mL microcentrifuge tube. The DNA can be further concentrated by alcohol precipitation if required. Dissolve DNA in TE buffer.

The DNA fragments are now in a concentrated and purified form and are ready for futher investigations.

B. Sucrose gradient purification

Centrifugation of fragmented donor DNA through sucrose gradients produced the purest DNA for cloning with an absence of ligation and/or packaging inhibitors.

1) Prepare two 38 mL sucrose (10-40% w/v) gradients in ultracentrifuge tubes using manufactured gradient-maker. The sucrose solutions are made in a buffer containing 1 M NaCl, 20 mM Tris HCl (pH 8.0) and 5 mM EDTA.

2) Load from 250-500 µg of partially digested donor DNA in a volume of ≤500 µL TE buffer onto each gradient. Overloading the gradient can lead to poor separation of fragments. Centrifuge samples at 26 000 rpm for 24 hours at 15°C.
3) Puncture bottom of tube with a 21 gauge needle and collect 0.5 mL fractions.
4) Remove an aliquot from every third fraction and analyze by electrophoresis through a 0.4% agarose gel with appropriate size markers. Photograph gel. Locate and pool the fractions containing the correct size fragments.
5) Dialyze the pooled fractions against 2-3 changes of 4 L TE buffer at 4°C for 16-24 hours.
6) Proceed to NACS column purification.

C. NACS column purification (from BRL NACS Prepac instruction manual)

NACS is an ion exchange resin that quantitatively binds nucleic acids in low salt (0.1-0.5 M NaCl) and releases them quantitatively in high salt (0.7-2.0 M NaCl). Each Prepac contains enough NACS-52 resin to bind up to 40 µg of A material. Optimal binding and recovery of nucleic acids on NACS Prepac is obtained when the sample does not contain organic solvents (chloroform, ethanol, ether, phenol), detergents (SDS, Triton, sarcosyl), proteins (BSA, enzymes), heavy metals, or polyvalent cations (polylysine, spermine). In general, optimal binding and recoveries of nucleic acids are obtained when a phenol extraction followed by ethanol precipitation is performed prior to NACS Prepac chromatography.

Mini-column method for NACS Prepac

1. Wash the resin 3 times with 2.0 M NaCl in TE buffer. This hydration is best accomplished by attaching the Prepac to the barrel of a 1 mL Pipetman. The solution is drawn up through the bottom of the column into the reservoir and then expelled by depressing the plunger. Flow through the NACS column can be bidirectional for the hydration procedure.
2. Wash the resin with Buffer C. The physical technique for this is the same as that used for hydration in step 1.
3. At this point it is usually more convenient to remove the Pipetman and clamp the Prepac onto a ring stand. The sample DNA (dissolved in Buffer C) is added to the top of the column and allowed to flow through by gravity flow in order to maximize binding. The column reservoir funnel may be attached to the top of the mini-column to increase the capacity above the resin to 6 mL.
4. Add 3-5 mL Buffer C to the column and allow it to flow by gravity or accelerate the flow with the Pipetman.
5. Elute the mini-column by gravity flow 3 times with 100-200 µL of Buffer D. Force out any remaining buffer with a Pipetman.
6. Add 1/10 volume 3 M sodium acetate and 2 volumes cold ethanol to precipitate the DNA. Wash in 70% ethanol. Resuspend in TE buffer.

D. Spermidine Precipitation of DNA

1. Adjust DNA solution to concentration of 100 µg/mL or less with TE buffer.
2. Add KCl to a final concentration of 100 mM and spermidine to a final concentration of 10 mM.
3. Incubate for 30 minutes on ice.
4. Centrifuge at 10 000 rpm for 10 minutes. Remove supernatant.
5. Wash pellet with solution containing 75% ethanol, 300 mM sodium acetate and 10 mM magnesium acetate. Agitate every 15 minutes briefly in vortex mixer for 1-4 hours.
6. Centrifuge at 10 000 rpm for 10 minutes. Remove supernatant.
7. Wash pellet with 70% ethanol for 30 minutes on ice.
8. Centrifuge at 10 000 rpm for 5 minutes. Dry briefly.
9. Resuspend pellet in TE buffer.

Appendix II: Media and buffer formulations

L broth	tryptone	10 g
	yeast extract	5 g
	NaCl	10 g
	water to	1 L

Adjust to pH 7.2, add $MgSO_4$ to 10 mM, and add 10 mL of 20% maltose per litre for growth of λ and its derivatives.

L broth agar	as L broth with addition of agar	15 g
L broth top agar	as L broth with addition of agar or agarose	6 g
TB broth	tryptone	10 g
	NaCl	5 g
	water to	1 L

After autoclaving add 10 mL of 1 M $MgSO_4$ and 10 mL of 20% maltose.

TB broth agar	as TB broth with addition of agar	15 g
TB broth top agar	as TB broth with addition of agar or agarose	6 g
CTAB extraction buffer (cetyltrimethylammonium-bromide)	Tris HCl (pH 8.0)	50 mM
	EDTA	5 mM
	Sorbitol	0.35 M
	BSA(Sigma A4503)	0.1%
	PEG 4000	10%
	Spermine	0.1%
	Spermidine	0.1%
	just prior to use, add β-mercaptoethanol	0.1%
CTAB wash buffer	Tris HCl (pH 8.0)	50 mM
	EDTA	25 mM
	Sorbitol	0.35 M
	just prior to use, add β-mercaptoethanol	0.1%
Mettler homogenizing buffer	sarcosyl	1%
	sucrose	0.25 M
	NaCl	50 mM
	EDTA	20 mM
	Tris HCl (pH 8.0)	50 mM
10X Sau3A buffer	Tris HCl (pH 7.4)	20 mM
	$MgCl_2$	5 mM
	KCl	50 mM
10X MboI buffer	Tris HCl (pH 8.0)	50 mM
	$MgCl_2$	10 mM
	NaCl	50 mM

50X TAE buffer	Tris base (pH 8.0)	242 g
	glacial acetic acid	57 mL
	EDTA (0.25M)	200 mL
Electrophoresis gel	1X TAE buffer	100 mL
	*agarose	0.4%
	ethidium bromide (10 mg/mL)	4 µL
	*or adjust to required concentration	
10X CIP buffer	Tris HCl (pH 9.0)	0.5 M
	$MgCl_2$	10 mM
	ZnCl	1 mM
	spermidine	10 mM
10X STE	Tris HCl (pH 8.0)	100 mM
	NaCl	1 M
	EDTA	10 mM
5X ligase buffer	Tris HCl (pH 7.8)	250 mM
	$MgCl_2$	50 mM
	DTT	100 mM
	ATP	5 mM
	BSA	250 µg/µL

Use fresh ligase buffer. Multiple freeze/thaws will decrease efficiency of ligation and thus the titer of packaging.

lambda diluent	Tris HCl (pH 7.5)	10 mM
	$MgCl_2$	10 mM
	EDTA	0.1 mM
phage dilution buffer (SM or SMC buffer)	Tris HCl (pH 7.5)	50 mM
	NaCl	100 mM
	$MgSO_4$	10 mM
	gelatin	0.2%
TE buffer	Tris HCl (pH 8.0)	10 mM
	EDTA (pH 8.0)	1 mM
Low Salt buffer	NaCl	0.2 M
	Tris HCl (pH 7.4)	20 mM
	EDTA	1 mM
High Salt buffer	NaCl	1 M
	Tris HCl (pH 7.5)	20 mM
	EDTA	1 mM
Buffer C	NaCl	0.5 M
	Tris HCl (pH 8.0)	10 mM
	EDTA	1 mM
Buffer D	NaCl	2 M
	Tris HCl (pH 8.0)	10 mM
	EDTA	1 mM

Host Strains	WA802 (mcrA$^-$ and mcrB$^-$) LE392 for control λ DNA ER1647 ER1648
Control DNA	ligated λ cI857 Sam7

Appendix III: Number recombinants required to represent genome of various species

Species	DNA amount (pg/nucleus)	DNA amount (bp/haploid genome)	Number recombinants for representative library
Juniperus virginiana	19.3*	9.75×10^9	2.61×10^6
Larix laricina	19.0*	9.5×10^9	2.57×10^6
Picea abies	29.6*	14.8×10^9	4.00×10^6
Picea glauca	17.0*	8.5×10^9	2.30×10^6
Picea mariana	22.2*	11.1×10^9	3.00×10^6
Pinus banksiana	29.8*	14.9×10^9	4.04×10^6
Pinus contorta	24.4*	12.2×10^9	3.30×10^6
Pinus lambertiana	35.2*	17.6×10^9	4.77×10^6
Pinus radiata	22.3*	11.2×10^9	3.03×10^6
Pinus resinosa	41.0*	20.5×10^9	5.55×10^6
Pinus rigida	17.8*	8.9×10^9	2.41×10^6
Pinus strobus	38.8*	19.4×10^9	5.26×10^6
Pinus sylvestris	27.6*	13.8×10^9	3.74×10^6
Thuja occidentalis	12.6*	6.3×10^9	1.71×10^6
Tsuga canadensis	27.0*	13.5×10^9	3.66×10^6
Nicotiana tabacum (NT-1)†	7.8	3.9×10^9	1.06×10^5
Arabidopsis thaliana†	0.16	8.0×10^7	2.17×10^4
Human+	6.6	3.3×10^9	7.60×10^5
Escherichia coli+	0.0084	4.2×10^6	9.20×10^2

*Miksche and Dhillon 1981.
†Rogers and Bendich 1985.
+Kaiser and Murray 1985.

Appendix IV: 5-methylcytosine in Plant DNA (Gruenbaum et al. 1981)

DNA-Sequence	% methylation
C-G	82
C-A	19
C-T	19
C-C	7
C-A-G	>80
C-T-G	>80
C-A-T	<4